Assessing Virtual Assistant Capabilities with Italian Dysarthric Speech

Assessing Virtual Assistant Capabilities with Italian Dysarthric Speech

Fabio Ballati
Politecnico di Torino
Torino, Italy
fabio.ballati@polito.it

Fulvio Corno
Politecnico di Torino
Torino, Italy
fulvio.corno@polito.it

Luigi De Russis
Politecnico di Torino
Torino, Italy
luigi.derussis@polito.it

ABSTRACT

The usage of smartphone-based virtual assistants (e.g., Siri or Google Assistant) is growing, and their spread was most possible by the increasing capabilities of natural language processing, and generally has a positive impact on device accessibility, e.g., for people with disabilities. However, people with dysarthria or other speech impairments may be unable to use these virtual assistants with proficiency. This paper investigates to which extent people with ALS-induced dysarthria can be understood and get consistent answers by three widely used smartphone-based assistants, namely Siri, Google Assistant, and Cortana. In particular, we focus on the recognition of Italian dysarthric speech, to study the behavior of the virtual assistants with this specific population for which there are no relevant studies available. We collected and recorded suitable speech samples from people with dysarthria in a dedicated center of the Molinette hospital, in Turin, Italy. Starting from those recordings, the differences between such assistants, in terms of speech recognition and consistency in answer, are investigated and discussed. Results highlight different performance among the virtual assistants. For speech recognition, Google Assistant is the most promising, with around 25% of word error rate per sentence. Consistency in answer, instead, sees Siri and Google Assistant provide coherent answers around 60% of times.

Author Keywords

Automatic Speech Recognition; Conversational Assistant; Dysarthria; Speech Impairment; Accessibility.

ACM Classification Keywords

• **Human-centered computing~Natural language interfaces** • **Human-centered computing~Empirical studies in accessibility**.

INTRODUCTION

In the first half of 2017, 42% of U.S. smartphone owners used AI-based personal assistants an average of 10 times per month. That is 71 million people with 710 million AI-based interactions. These are people using Siri, Google Assistant, Cortana, and other virtual assistants for nearly one billion hours per month [11]. The virtual assistants landscape has changed significantly over the past seven years. Siri, one of the earliest mobile personal assistants, was integrated into the iPhone in October 2011; Microsoft's Cortana debuted three years later. Similarly, Google has introduced a series of apps, including Google Now (released in 2012), Allo (2016), and Google Assistant (February 2017). By using speech as the primary input, virtual assistants can bypass or minimize the more "conventional" input methods (i.e., keyboard, mouse, and touch), thus making voice-controlled devices useful and accessible. However, while persons with motor disabilities may benefit from these virtual assistants, those with cognitive, sensory, or speech disorders may be unable to fully use them. For example, Bigham et al. [2] demonstrated that Google's speech recognition system does not work well for people who are deaf or hard of hearing, and they expected that recognizing deaf speech will remain challenging for both automatic and human-powered approaches.

This paper investigates to which extent people with speech impairments can use and be understood by the three most common smartphone-based virtual assistants. We focus on people with dysarthria a motor speech disorder characterized by poor articulation of phonemes that makes it difficult to pronounce words. In particular, we focus on people with amyotrophic lateral sclerosis (ALS) induced dysarthria whose intelligibility of speech, evaluated with the "Speech" category of ALS Functional Rating Scale (FRS-r) [3], was 'detectable speech disturbance' (value 3) or 'intelligible with repeating' (value 2). In addition, we assess virtual assistants with people that are native Italian speakers, thus choosing to focus on the recognition of Italian dysarthric speech. This because, at the best of our knowledge, there are no available studies on the Italian speech recognition and the related answers made by virtual assistants. We investigated the interaction of people with dysarthria with the three most used virtual assistants for mobile devices: Apple's Siri, Google Assistant, and Microsoft's Cortana. The goals of the paper are both to define whether people with moderate dysarthria could be understood by the three virtual assistants (*question*

comprehension, *QC*) and to investigate which assistant provides the most coherent answer (*consistency in answer*, *CiA*) when the recognized speech is correct or partially wrong. *QC* represents the similarity between the expected transcription of a voice request and the transcribed output of each assistants. *CiA*, instead, indicates the appropriateness of the assistants' responses, i.e., it represents the percentage of times that an assistant provided a given type of answer to a voice query.

For these purposes, we designed a specific methodology and we recorded 34 sentences from patients with Amyotrophic Lateral Sclerosis (ALS) at the Otolaryngology department of the Molinette hospital, in Turin, Italy. In fact, people with ALS are often affected by dysarthria since it typically results from a neurological injury of the motor component of the motor-speech system. A secondary but relevant contribution of the paper is the availability of a consistent set of dysarthric Italian spoken sentences, that will be published on our website and that might benefit other researchers, too. By using this set of dysarthric Italian spoken sentences, it will be possible for other researchers and practitioners to replicate the experiment, and optionally expand it with other sentences. In our study, we played each speech recording to every virtual assistant, separately, and we analyzed both the given transcription and the assistants' answers. We assessed the accuracy in transcription of the dysarthric sentences, to define the *QC* of the assistants. Later, we focused on the *CiA*, to find out whether the three assistants give coherent answers. Results show that the three virtual assistants have different performance for both *QC* and *CiA*. In terms of *QC*, the average Word Error Rate (WER) for transcribed questions varies from Google Assistant (24.88%) to Cortana (39.39%), to Siri (70.89%). Considering *CiA*, the percentage of coherent answer (calculated for the correctly recognized questions) is higher for Siri and Google Assistant (around 60%) than for Cortana (25%). Among the three assistants, Google Assistant is the one that performs better when used by people with dysarthria.

To summarize, the main contributions of this paper are:

1. the proposed methodology, which aims to allow replication and extension of the experiment,
2. the collected dataset, which will be available to other researchers, and
3. the fact that the results are applicable to a specific combination of a minor language (Italian) and a well-defined disability (ALS-induced dysarthria).

RELATED WORKS AND BACKGROUND

Speech technology in general, and automatic speech recognition (ASR) in particular, are not new for people with disabilities. They have been used to increase accessibility in mainstream operating systems since decades, as an alternative method to compose documents through dictation systems or to control the computer and, recently, to control the smartphone. Similarly, speech recognition as an input to electronic assistive technology was investigated both in general and for dysarthria. Hawley [7] presents an early overview, based on a literature review and clinical observations, upon the suitability and performance of speech recognition for computer access by people with disabilities, including people with dysarthria. He reports that, given adequate time, training, and support, commercial ASR systems for PCs are often appropriate for people with no, mild, or moderate speech impairments. People with dysarthria achieve lower recognition rates, but speech recognition can be still a useful input method for some individuals. Conversely, Hawley discovers that speech as a mean of controlling electronic devices such as smartphones and appliances is more troublesome, especially for dysarthric speech. To overcome this kind of issues, researchers investigated new methods and proposed dedicated ASR systems for dysarthria, e.g., by using ergodic hidden Markov models [12] or articulatory dynamic Bayes networks [13].

Specific HCI research in the domain of technology for people with speech impairments is, instead, still quite limited [4] for English language, and it is totally absent for other less spoken languages, like Italian. Sears et al. [14] offer an overview of HCI research for people with "significant speech and physical impairments", by focusing on communication aids. More recently, Derboven et al. [4] describe the design of ALADIN, a self-learning speech recognition system for people with physical disabilities, many of whom also have speech impairments. ALADIN is designed to allow users to use their own specific words and sentences, adapting itself to the speech characteristics of the user.

Finally, a few works explore usability and accessibility issues of virtual assistants. Lopez et al. [8] present a usability evaluation of some speech-based virtual assistants (i.e., Alexa, Siri, Cortana, and Google Assistant) and highlight that there is still a lot to do to improve the usability of these systems. Glasser et al. [6], instead, focus on the issues that may arise from the usage of two virtual assistants by people who are deaf and hard of hearing. Bigham et al. [2] propose two technical approaches for enabling deaf people to provide input to those assistants, i.e., human computation workflows for understanding speech and mobile interfaces that can be instructed to speak on the user's behalf. Ballati et al. [1], instead, investigated the interaction of English dysarthric speech data with three widely used virtual assistants, included in several standalone and mobile devices (Apple's Siri, Google Assistant, and Amazon Alexa). Similar to both the work of Glasser et al. and Ballati et al., we focus on the issues that may arise from the usage of virtual assistants, but we are specifically interested in Italian dysarthric speech and in the evaluation of the current behavior of the most used smartphone-based virtual assistants: Google Assistant, Siri, and Cortana.

Dysarthric Speech

Dysarthric speech is the speech produced by people with dysarthria. Dysarthria can result from congenital conditions, or it can be acquired at any age as the result of neurologic injury, disease, or disorder. Dysarthria refers to a group of neurogenic speech disorders characterized by "abnormalities in the strength, speed, range, steadiness, tone, or accuracy of movements required for breathing, phonatory, resonatory, articulatory, or prosodic aspects of speech production" [5]. These abnormalities are due to one or more sensorimotor problems – including weakness or paralysis, incoordination, involuntary movements, or excessive, reduced, or variable muscle tone. These sensorimotor problems distort motor commands to the vocal articulators, thus resulting in atypical and relatively unintelligible speech in most cases. Dysarthric speech may be characterized by a slurred, nasal-sounding or breathy speech, an excessively loud or quiet speech, problems speaking in a regular rhythm, with frequent hesitations, and monotone speech. As a consequence of these problems, a person with dysarthria may be difficult to understand and, in some cases, she may only be able to produce very short phrases, single words, or no intelligible speech at all. Consequently, enabling modern ASR to effectively understand dysarthric speech is a major need, both for virtual assistants and for computers, since other physical impairments often associated with dysarthria can render other forms of input, such as keyboards or touch screens, especially difficult. To provide an estimated measure of the people with ALS-induced dysarthria, we started from the data about ALS. It is generally estimated there are around 450,000 people living with ALS worldwide [15]. Dysarthria occurs in more than 80% of ALS patients and may cause major disability [16].

STUDY

In this section, we first discuss how we collected and recorded speech samples. Then, we show the study procedure for the speech recognition, and we conclude by illustrating the evaluation criteria used for *QC* and *CiA*.

Data Collection

To explore the issues of understanding dysarthric speech by contemporary virtual assistants, we recorded 34 Italian sentences from eight people. Participants were all native speakers, with ALS-induced dysarthria, restricting to two speech intelligibility categories. The sentences are a mix to recommended questions from Amazon Echo [9] and Google Home [10], modified to include all the phonemes of the Italian language. The recommended questions for Siri and Cortana were considered, but not included, because they are very similar, and do not add qualitatively different sentence types. Table 5 reported the full list of sentences with their English translations. The participants involved in this phase are some of the patients of the phoniatric and logopedic clinic in the Otolaryngology department of the Molinette hospital, in Turin, Italy. The clinic is managed by a phonatrist and a speech therapist expert in ALS. The clinic has the purpose of contributing to the diagnosis, treatment, and monitoring of swallowing disorders (dysphagia) and articulation of language (e.g., dysarthria).

The evaluation for the speech intelligibility of participants was made by using the "Speech" category of Amyotrophic Lateral Sclerosis Functional Rating Scale (*ALS FRS-r*) [3], which identifies the severity of the disease and is subdivided for skill categories. We select people whose intelligibility of speech, evaluated by the speech therapist, was *detectable speech disturbance* (6 people) or *intelligible with repeating* (the remaining two persons). The speech therapist in collaboration with the phoniatrist also help define the types of dysarthria for the people involved in this study. The eight patients have three types of dysarthria, as evaluated according to the Duffy classification [5]:

1. *Flaccid*, associated with disorders of the lower motor neuron system and/or muscle.
2. *Spastic*, associated with bilateral disorders of the upper motor neuron system.
3. *Unilateral upper motor neuron*, associated with unilateral disorders of the upper motor neuron system.

Table 1 shows the data about sex (M stands for Male and F stands for Female), age, type of dysarthria and intelligibility of speech, for each patient.

User	Age	Dysarthria	ALS FRS-r
M1	77	Flaccid	3 (Detectable speech disturbance)
M2	80	Spastic	2 (Intelligible with repeating)
M3	64	Spastic	3 (Detectable speech disturbance)
M4	72	Unilateral Upper Motor Neuron	3 (Detectable speech disturbance)
F1	83	Flaccid	3 (Detectable speech disturbance)
F2	72	Flaccid	2 (Intelligible with repeating)
F3	67	Mixed (Spastic and Flaccid)	3 (Detectable speech disturbance)
F4	71	Unilateral Upper Motor Neuron	3 (Detectable speech disturbance)

Table 1. Characteristics of the 8 users with dysarthria involved in the experiment

Each participant voluntarily accepted to participate in the study and agreed to sign the informed consent for the processing of personal data and his or her recorded voice. The recordings took place in the same clinic where they usually have their phoniatric examination, and the accompanying family members were present, too. Data was collected between January 2018 and March 2018. For the recording process, we used a three phases protocol. First, we explained the purpose and the goals of the study, and the method adopted for the recordings. Second, the participant read each sentence from an A4 sheet of paper (one sheet for each sentence), located in front of the reader, at a suitable distance. We recorded their voice, while they read each sentence. The recordings were taken with a smartphone, located at a distance of 30-40 centimeters, i.e., the distance recommended by the audio recording application in use. The data consists of audio files recorded using the *Recforge II* application on a Samsung A5 2017 smartphone. We chose to use a smartphone to record the audio samples starting from read sentences to ensure that all participants would speak exactly the same sentences, thus obtain a consistent dataset, more easily replicable and extensible. Furthermore, smartphone is the main and preferred input channel for the smartphone-based virtual assistants we are evaluating. Finally, we gave to each participant a questionnaire to understand whether the read phrases could be useful in their everyday life. In general, all the participants said that the sentences are useful in everyday life.

Study Procedure

For the speech recognition procedure, we use the audio files of the 34 sentences reported in previous section. The experiment took place in a quiet room of our university. The speech samples were played on a laptop connected to an external high quality speaker. Each sentence was played for each virtual assistant, separately, and the results of the operation (i.e., recognized request and related response) were noted down by the experimenter. The virtual assistants were run on dedicated smartphones: an iPhone 7 (iOS 11.2) was used for Siri, a Samsung A5 2017 (Android 8.1) for Google Assistant, and a Lumia 910 (Windows 10 Mobile) was used for Cortana. We evaluated the accuracy of the speech-to-text recognition process adopted in terms of *question comprehension* (both quantitatively and qualitatively), and the related responses in terms of *consistency in answer*.

Measures

We evaluated the accuracy of the speech-to-text recognition process adopted by the virtual assistants using two different criteria of question comprehension (QC): *quantitative QC* and *qualitative QC*. All the assistants, indeed, provide the user with a transcription of the received command. The evaluation of the *QC* is, therefore, given by the similarity between the expected transcription (i.e., the 34 original sentences read by users with dysarthria) and the transcribed output of the assistants.

For *quantitative QC*, we computed the Word Error Rate (WER) between the expected transcription and the actual transcribed sentence. The WER is defined as: WER = (S + I + D) / (S + I + C) (I=insertion, D=deletion, S=substitution, C=correct).

For *qualitative QC* we analyzed each transcribed sentence and classified it according to five categories:

a) Correct questions, i.e., the transcribed sentences are equal to the expected transcription.
b) Questions with the same semantic meaning: i.e., transcribed sentences are equal to the expected transcriptions in terms of semantics, but not literally identical. For example, the sentences “Set the house temperature to 22 degrees” and “Set in the house temperature 22 degrees” are not equal, but they have the same semantic meaning.
c) Incomplete questions: i.e., sentences only partially transcribed, where the transcribed portion of the sentence is correct.
d) Wrong questions: i.e., sentences semantically and syntactically different with respect to the expected sentences.
e) Not recognized questions: i.e., sentences not recognized at all by a virtual assistant.

To avoid arbitrary choices, two researchers assessed each transcribed sentence, individually. In cases of differing evaluations, they discussed to reach agreement.

Even if the accuracy of the speech-to-text recognition process is not perfect, virtual assistants may leverage the context or some specific recognized keywords to provide a suitable response. We only evaluated the answer given by the three virtual assistants in cases of properly transcribed *questions*, i.e., sentences belonging to the previous a) and b) categories. *CiA* is, instead, an indicator of the appropriateness of the assistants’ responses, given as the number and percentage of times that the three assistants provided a certain type of answer to the user’s queries. We classified the responses in 3 classes:

1. Coherent answers: i.e., correct responses or responses logically consistent with the asked question. E.g., if the question is “Which is the nearest supermarket?” and the assistant says, “Here the result for ‘supermarket’ within 3 km” and shows the map with the exact position of nearby supermarkets.
2. Incoherent answers: i.e., responses logically incorrect to the questions. E.g., if the question is “How much does a flight to Rome cost?” and the incoherent answer is “Rome beat Crotone in the last soccer match”.
3. Default answers: i.e., responses that a assistant provides by default, when it is not able to fully understand the request nor to extract any context. E.g. if the question is “How much is 42 multiplied

by 76 divided by 3?" and the default answer is "Here is what I found looking for '' How much is 42 multiplied by 76 divided by 3?" on the web.

RESULTS

The result for *quantitative QC, qualitative QC, and CiA* are illustrated below.

User	Google Assistant	Cortana	Siri
M1	0.00% (0%)	24.25% (23.59%)	46.14% (36.37%)
M2	63.92% (37.9%)	67.89% (30.10%)	96.08% (8.67%)
M3	13.19% (20.19%)	24.01% (31.60%)	30.36% (30.05%)
M4	15.85% (25.54%)	26.74% (28.83%)	56.24% (30.55%)
F1	14.09% (23.23%)	19.04% (24.03%)	70.87% (29.2%)
F2	30.02% (36.65%)	34.74% (31.81%)	98.86% (6.76%)
F3	20.54% (25.58%)	48.99% (39.32%)	74.38% (29.29%)
F4	41.41% (36.82)	69.45% (32.37%)	94.21% (17.82%)

Table 2. Average word error rate for each user by using each virtual assistant. (SD in parenthesis).

Quantitative QC

For what concerns the *quantitative QC*, table 2 shows the average WER for each user and for each virtual assistant. Also, the results show that the WER is highly dependent upon the user. In fact, considering all assistants, the average WER for users M1 and M3 is around 20%. Conversely, for M2 the average WER is around 75% and it is little less than 70% for F4. The average WER for Google Assistant, considering all questions and users, is lower than the average WER for Cortana (WER: 24.88%, SD: 33.61% for Google vs. WER: 39.39%, SD: 35.65% for Microsoft). Furthermore, the average WER for Siri (70.89%, SD: 34.70%) is sharply higher compared with both Google Assistant and Cortana. Figure 1 shows the distribution of the WER values for each platform. To investigate whether the differences in the measures were statistically significant, we analyzed the effect of the independent variable (Google Assistant vs. Cortana vs. Siri) over the dependent variables (WER) with a one-way repeated measure ANOVA. The test shows that there is a significant effect of the used virtual assistants ($F(2,14) = 30.06$, $p < .01$). Looking at individual users, for each participant results confirm the same trend, worsening as we move from Google Assistant to Cortana and to Siri.

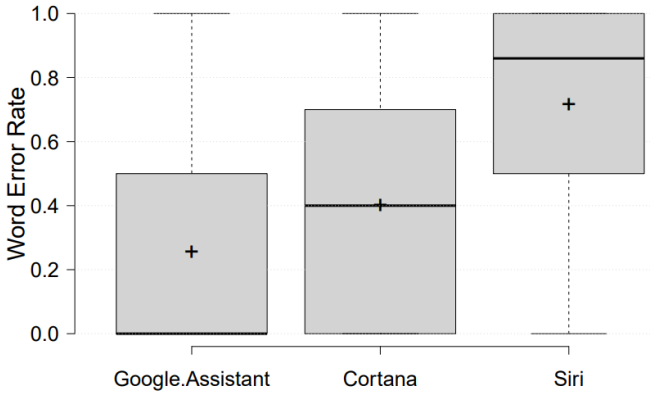

Figure 1. WER distribution for each assistant

Qualitative QC

For what concerns the *qualitative QC*, the behavior of the virtual assistants is considerably different among them and similar to *quantitative QC*. Table 3 presents, for each assistant, the quantities and percentages of recognized questions, by categories.

	Properly transcribed questions				
	a) Correct	b) Same semantic meaning	c) Wrong	d) Incomplete	e) Not recognized
Google Assistant (272)	135 (49.63%)	39 (14.33%)	58 (21.32%)	39 (14.33%)	1 (0.37%)
Cortana (272)	85 (31.25%)	23 (8.45%)	141 (51.83%)	20 (7.35%)	3 (1.10%)
Siri (272)	36 (13.23%)	7 (2.58%)	149 (54.78%)	32 (11.76%)	48 (17.65%)

Table 3. Quantity and percentage of question recognized by conversational assistants, by categories

According to Table 3, about 62% of the sentences transcribed by Google Assistant are *properly transcribed questions*, while the *wrong questions* are the 21.32% of total sentences by Google Assistant. Looking at Siri and Cortana, Table 3 shows that, for these two virtual assistants, the number of *wrong questions* is more than twice than Google Assistant's. Conversely, Siri and Cortana have different results for what concerns *properly transcribed questions* and other remaining questions (i.e., questions that do not belong to the previous two categories). In case of *properly transcribed questions*, 40% of sentences recognized by Cortana belong to *properly transcribed questions* and 15% by Siri, only. For what concerns the remaining questions, Table 3 highlights that, for Siri, many sentences are *Not recognized* questions. For Cortana, most of the remaining questions belongs, instead, to the *Incomplete questions* category (the same is for Google Assistant).

CiA (*Consistency in Answer*)

Finally, also for *CiA* the behavior of the three virtual assistants is different. The answers provided by Google Assistant and Siri are in most cases coherent, and in a small number of cases the sentences retrieved from the two virtual assistants are the default answers. Table 4 presents, for each assistant, the quantity and percentage of coherent answers, default answers, and incorrect answers. Table 4 also shows that Siri and Google Assistant answered consistently between 50 and 60% of cases, and around 30% - 40% are the default answers. Instead, in most cases, Cortana gives the default answers (75.93%), while the coherent answer is given in a few cases (24.07%), only.

Properly transcribed questions (a + b)	**Coherent answer**	**Default answer**	**Incorrect Answer**
Google Assistant (174)	94 (54.02%)	78 (44.83%)	2 (1.15%)
Cortana (108)	26 (24.07%)	82 (75.93%)	0 (0%)
Siri (43)	26 (60.47%)	13 (30.23%)	4 (9.30%)

Table 4. Quantity and percentage of coherent answers, default answers and incorrect answers

DISCUSSION

Starting from the results about *QC* and *CiA*, this discussion aims at bringing out the different behavior of the three virtual assistants.

The result for *quantitative QC* shows that the behavior of the three virtual assistants is significantly different, but also underlines that the accuracy in transcription is strictly

related to the specific user, for all three virtual assistants. Particularly, overall and for each user, Google Assistant provides the best results (i.e., the lowest WER). On the other hand, Siri has the higher WER both overall and for each user, even if for users without speech impairments the WER for Siri is around 5%, like other industry-leading voice recognition system [11]. In terms of WER, Cortana is halfway between Siri and Google Assistant. Cortana, for each user, is able to recognize the question with an average WER which is in between the other two virtual assistants, in some cases very close to Google Assistant results (M2, F1, and F2) and in other cases close to the Siri result (M3). The result for user M1 clearly shows the differences between the three virtual assistants. The WER for Google Assistant is 0%, the same value for Cortana is around 25%, and the value for Siri is 46.14%. In this case, M1 could use Google Assistant without any problems, but the same does not apply to Cortana and Siri.

Result for *qualitative QC* confirms what has emerged from *quantitative QC* about the considerably different behavior of the three virtual assistants. Considering all users, in fact, Table 3 shows that Google Assistant has the highest number of *properly transcribed questions* (174 out of 272). This value for Cortana is 108 out of 272, compared to 43 out of 272 for Siri, only. The results for Google in terms of *properly transcribed questions* are the best compared with the other two virtual assistants. In addition, in many cases the questions transcribed by Google Assistant are in the *Incomplete questions* category. For this category, Google Assistant correctly transcribed the first part of speech sample and, in many cases, it stopped listening due to an interval of silence, quite usual in dysarthric speech. Indeed, talking is very tiring for people with dysarthria and the pronunciation of the sentences is slower than for people without speech impairments. Therefore, the recognized sentence is cataloged as an incomplete question, but potentially if Google Assistant continued listening the speech sample, the transcription could have been correct, and the number of *properly transcribed sentences* might further increase. On the other hand, Siri has the highest number of *Wrong questions*. In addition, Siri not recognize at all a large number of questions. Overall, Siri badly recognizes 197 out of 272 questions, more than 70%. As in *quantitative QC*, the result for Microsoft are between Google Assistant and Siri.

The *QC* analysis points out that the only virtual assistant currently usable by people with dysarthria is Google Assistant, but the usability is strictly related to how the voice of the user and their vocal apparatus are affected from dysarthria. Instead, question comprehension appears not to be directly related to the different types of dysarthria.

For *CiA*, the results show the different behavior of the three virtual assistants. Both Google Assistant and Siri have good percentage of coherent answers related to the *properly transcribed questions*. On one hand, Siri has the best percentage value for coherent answers (60%) corresponding to 26 correct answers out of 46 *properly transcribed*

questions. On the other hand, Google Assistant has a similar value (54%), but it is particularly significant since it is obtained from 94 coherent answers out of 174 starting *properly transcribed questions*. The remaining (and minor) set of answers, mainly, are the default answers both for Siri and for Google Assistant. Conversely, Cortana gives the default answer to most of the questions (about 75%). Cortana never provides incoherent answers, probably because in cases of doubt it gives the default answer. Unfortunately, we acknowledge that due to the high Siri recognition errors, the number of properly transcribed questions is quite small (i.e., 43) in this case. Finally, in several cases of incomplete question transcription, Siri gives an unsuitable answer, not pleasant to read. In fact, some incomplete sentences occur when the virtual assistants stop listening due to an interval of silence in the speech sample. So, in addition to stopping the speech recognition process before the person with dysarthria completes her question, Siri replies something like "You'd better consult a crystal ball".

CiA points out that the behavior of Google Assistant and Siri is similar. These virtual assistants are useful to obtain a precise answer to specific questions. On the contrary, Cortana usually gives the default answer (i.e., it searches on the Web).

Study Limitations

We would like to acknowledge that this study presents three main limitations. First, the relatively low number of subjects involved in the experiment. This is due to the limited number of persons with moderate dysarthria, inducted by ALS, living in our region. Second, we recorded the voice with a mobile phone because the smartphone is the normal device which is really usable by patients in their everyday life and is the platform where voice assistants are expected to be run. For this reason, the quality of the audio files may be not particularly excellent compared withaudio recordedwith professional microphones. However, this choice will allow an easier replication and extension of the study. The third limitation stems from the choice of playing sentences from a speaker instead of by a human being. This was inevitable since, in this work, we chose to compare the three assistants with exactly the same audio sample, and they run on different platforms. We should also notice, however, that we do not have any evidence that this choice negatively impacted the results of the evaluation.

CONCLUSION AND FUTURE WORK

Smartphone-based virtual assistants enable people to more easily control their smartphones and to quicken the actions they already take on their phone and other devices. However, a consequence of these devices embracing voice control is that people with dysarthria or other speech impairments may be unable to use them with profit.

In this paper, we investigate the interaction of people with dysarthria induced by ALS, with three of the most used virtual assistants for mobile devices (Apple's Siri, Google Assistant, and Microsoft's Cortana). We are interested in defining whether Italian speech of people with moderate dysarthria could be understood by the three virtual assistants. Moreover, we investigated the *consistency in answer (CiA)* to understand which of the three virtual assistants provides the most consistent and useful replies, even in case of partial question understanding. For these purposes, we used 34 sentences that we have recorded from Italian dysarthric speakers.

The results show that the behavior of the three virtual assistants is considerably different and that the accuracy in transcription is strictly dependent on the user, for all three virtual assistants. On one hand, about *question comprehension* (*QC)*, people with moderate dysarthria could be quite easily understood by Google Assistant, but not perfectly. On the other hand, about *CiA*, both Google Assistant and Siri are useful to obtain a precise answer to a specific question, instead Cortana usually gives the default answer.

Future work will include the publication of the set of dysarthric Italian speech data. Furthermore, we will record the same speech samples from users without speech impairments, to perform a comparison with a similarly-sized control group. Finally, we will use the outcome of this evaluation as a starting point to improve the accessibility and the recognition capabilities of such assistants.

REFERENCES

1. Fabio Ballati, Fulvio Corno, Luigi De Russis. In Press. "Hey Siri, do you understand me?": Virtual Assistants and Dysarthria. In *Workshop Proceedings of the 14th International Conference on Intelligent Environments (IE 2018)*. IOS Press, pages 10.
2. Jeffrey P. Bigham, Raja Kushalnagar, Ting-Hao Kenneth Huang, Juan Pablo Flores, and Saiph Savage. 2017. On how deaf people might use speech to control devices. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '17), 383–384. https://doi.org/10.1145/3132525.3134821.
3. Jesse M. Cedarbaum, Nancy Stambler, Errol Malta. 1999. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*, 169, 1–2: 13-21. https://doi.org/10.1016/S0022-510X(99)00210-5.
4. Jan Derboven, Jonathan Huyghe, and Dirk De Grooff. 2014. Designing voice interaction for people with physical and speech impairments. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (NordiCHI '14), 217-226. https://doi.org/10.1145/2639189.2639252.
5. Joseph Duffy. 2012. Motor Speech Disorders, Substrates, Differential Diagnosis, and Management.

6. Abraham T. Glasser, Kesavan R. Kushalnagar, and Raja S. Kushalnagar. 2017. Feasibility of Using Automatic Speech Recognition with Voices of Deaf and Hard-of-Hearing Individuals. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '17), 373-374. https://doi.org/10.1145/3132525.3134819.
7. Mark S. Hawley. 2002. Speech recognition as an input to electronic assistive technology. *British Journal of Occupational Therapy* 65, 1: 15–20. https://doi.org/10.1177/030802260206500104.
8. Gustavo Lopez, Luis Quesada, and Luis A. Guerrero. 2018. Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. In *Proceedings of the AHFE International Conference on Human Factors and Systems Interaction* (AHFE 2017), 241–250. https://doi.org/10.1007/978-3-319-60366-7_23.
9. Taylor Martin, David Priest. 2017. The complete list of Alexa commands so far. Retrieved April 11, 2018 from https://www.cnet.com/how-to/amazon-echo-the-complete-list-of-alexa-commands/.
10. Taylor Martin, David Priest. 2017. The complete list of Google Home commands so far. Retrieved April 11,
11. Erin Myers. 2017. Speech Recognition Accuracy: Past, Present, Future. Retrieved April 11, 2018 from https://www.temi.com/blog/2017/10/06/speech-recognition-accuracy-history/.
12. Prasad D. Polur and Gerald E. Miller. 2006. Investigation of an HMM/ANN hybrid structure in pattern recognition application using cepstral analysis of dysarthric (distorted) speech signals. *Medical Engineering & Physics*, 28, 8: 741 – 748.
13. Frank Rudzicz. 2012. Using articulatory likelihoods in the recognition of dysarthric speech. *Speech Communication*, 54, 3: 430 – 444.
14. Andrew Sears and Mark Young. 2002. Physical disabilities and computing technologies: an analysis of impairments. *In The human-computer interaction handbook* 482-503.
15. Therapy Development Institute ALS. 2018. ALS Frequently Asked Questions. Retrieved June 13, 2018 from https://www.als.net/about-als-tdi/als-faq/.
16. Barbara Tomik and Roberto J. Guiloff. 2010. Dysarthria in amyotrophic lateral sclerosis: A review. Amyotrophic Lateral Sclerosis. 11, 1-2: 4-15. https://doi.org/10.3109/17482960802379004.

Italian Sentence	English Translation
Quando sarà la prossima partita della Juventus?	When will the next Juventus match be?
Tra quanto tempo passerà alla stazione di Torino Porta Susa il prossimo treno regionale per Chivasso, in arrivo dal lingotto?	How soon before will pass at the Torino Porta Susa station the next local train to Chivasso, arriving from Lingotto Station?
Oggi ho bisogno di prendere l'ombrello?	Do I need to take an umbrella, today?
Quale sarà la prossima partita della Seria A di calcio?	When will the next football match of Serie A?
Quante proteine ci sono in due uova?	How many proteins are there in two eggs?
Quali ingredienti servono per gli spaghetti alla bolognese?	What are the ingredients for spaghetti alla bolognese?
Aggiungi cipolla e pomodori alla mia lista della spesa.	Add onion and tomatoes to my shopping list
Quanto tempo ci vuole per arrivare dall'università alla stazione del treno?	How long does it take to get from the university to the train station?
Chi è l'attuale presidente della repubblica italiana?	Who is the current president of the Italian republic?
Quand'è il miglior momento per seminare i semi?	When is the best time to sow seeds?
Cosa ci sarà in tv questa sera?	What will there be on TV tonight?
Quando saranno i prossimi saldi di Benetton?	When will Benetton's next sales be?
Quanta energia sprechi?	How much energy do you waste?
Quanto fa 42 moltiplicato 76 diviso 3?	How much is 42 multiplied by 76 divided by 3?
Quanto tempo ci metto per andare a lavoro?	How long does it take me to go to work?
Quante calorie ci sono in una banana?	How many calories are there in a banana?
Qual è il supermercato più vicino?	Which is the nearest supermarket?
Quali sono le previsioni per il week end?	What are the forecasts for the weekend?
Imposta una sveglia alle 8.	Set an alarm at 8.
Chi era ministro italiano dell'istruzione nel 2005?	Who was the Italian education minister in 2005?
Come si dice casa in inglese?	How do you say home in English?
Che verso fa la zebra?	What is the cry of the zebra?
Che giorno è oggi?	What day is today?
Raccontami la trama del film "Il signore degli anelli".	Tell me the plot of the movie "The Lord of the Rings".
Qual è lo spelling della parola ciliegia?	What is the spelling of the word cherry?
Imposta la temperatura della casa a 22 gradi.	Set the house temperature to 22 degrees.
In quale giorno si festeggia la festa della repubblica in Italia?	On what day is the Republic Day in Italy celebrated?
Chi è l'attore protagonista del film "Una settimana da Dio"?	Who is the leading actor in the movie "Bruce Almighty"?
Quanto valgono attualmente 100 euro in dollari?	How much are currently worth 100 euros in dollars?
Quanto costa un volo andata e ritorno per Roma?	How much does a flight to Rome cost?
Quanto impiega il taxi per percorrere la strada di ritorno a casa?	How long does the taxi take to get back home?
Quali sono le regole di una partita di scopa?	What are the rules of the card game "Scopa"?
Lo sci è uno sport completo per l'allenamento di tutto il corpo?	Is skiing a complete sport for whole body training?
Qual è il numero telefonico della croce rossa?	What is the red cross telephone number?

Table 5. The 34 sentences read by people with dysarthria